Notes from ACRID teleconference, 22nd February 2011

Attending: Arif Shaon (STFC), Spiros Ventouras (STFC), Sarah Callaghan (STFC), Tim Osborn (UEA), Colin Harpham (UEA), Jeremy Tandy (Met Office)

 Apologies: Ag Stephens (STFC), Bryan Lawrence (STFC)

Next meeting: to be confirmed, but will be a face-to-face meeting held at STFC.

Actions:

- Arif – send latest versions of workflow documents to the list. Include CRU TS.
- Arif and Jeremy get together to work on the information model in person.
- Sarah – set up doodle poll for next meeting
- Arif – coordinate production of deliverable D2.2 due March 1st
- Colin - coordinate production of deliverable D2.1 due March 1st
- Sarah – coordinate production of deliverable D1.3 due March 1st

Previous Actions:

- Tim – put slides and notes from December meeting on project webpage – ongoing and update contact details - done.

Completed Actions:

- Sarah – send Jeremy and Ag draft workflows for CRUTEM and tree-ring data etc - done.
- Sarah – email Simon about deliverables – done.
- Set up a mailing list for the group – **Sarah - done**
- Sarah - set up another doodle for a telco next month - done.
- Sarah – email document format for mid-term report to team - done.

Issues:

- How do we reconcile the complexity of ISO 19115 with the O&M model and lighter touch we want for our information model?

Decisions:

- Main part of D2.2 is the information model.
- Focus on core concepts. Won't have fully comprehensive model by the end of this project – can be built on later.

Minutes and actions from the previous meeting
Project management update

Don't forget we've got three deliverables due on the 1st March (D1.3 Mid-term report, D2.1 Description of scientific workflows and D2.2 Information architecture).

D2.2 – includes information model we've all been working on. Arif started writing things up – model quite volatile at this stage. Need to freeze the information model and report on the current version. – Arif in charge.

D2.1 – mostly done. Just need to put them together in a structured format – Colin in charge.

CRU TS workflow isn't finalised – diagrams need text added.
JISC MRD workshop – Sarah and Arif going, Arif presenting. Colin and Tim will look at the agenda and decide if they want to attend.

Next meeting/teleconference

Next meeting have as a face-to-face meeting to really discuss the information model in a month or so (after 21st March), hosted at STFC.

WP2.2 – hardly any work done on data management infrastructure. Diagram sent around as basis for future work. Had some feedback. Information model is the key part of the deliverable so data management infrastructure diagram should be ok. Doesn't capture BADC dependency for CRU TS dataset.

Progress update from STFC

Spiros and Arif working on information model. Specialisation of O&M model. Generalisation of concepts captured by the workflows. CRU datasets are representations of observations made of certain features of interest. Currently on internal version 3.2.

Progress update from UEA

Working on CRU TEM dataset – trying to organise the way the data are held in a subversion repository. Allows UEA to exchange and compare datasets more easily with Met Office. Number of scripts put together to facilitate this and created introductory guidance. People at UEA starting to use these for real work.

Initial setting-up of repository for tree-ring and CRU TS started. Not being used yet, but hopefully soon.

Progress with talking to colleagues about capturing of metadata. Lot of metadata involved in processing of tree-ring data compared with CRU TEM. Want to be able to reproduce the output from the collected metadata. Software now records parameters and options at the top of the files. Metadata not yet associated with the version of the software.

Software products – tool or mechanism for capturing metadata. Need to understand the data management environment.

Need to capture information and display it in the right format. Presentation can be done via GeoTOD. Concentrate on capturing.

To recreate an analysis we need:
1. Raw tree ring measurements
2. Which version of tree-ring standardisation software was used
3. What options and settings of the software were used (currently captured)

Tree ring chronology is published as a different thing from the processed data. Don't need to be published at the same time, but can be part of the same workflow.

Model will be able to define all the observations. Should we publish the raw measurements separately from any products derived from it (business decision)? No technical impediment to publishing these separately.

Discussions

Related resource – trying to capture anything related to an observation outside of the domain being described by the model, e.g. publication, CRU TEM v 2... Can discard it later on if needed.

Lineage information in the ISO 19115 model. Might be a bit too complex when it comes to populating the model.

Result type has to be of discrete grid point coverage. CSML grid series observation type doesn't have related resource.

Semantics for process is to describe a type or class of process. Different to how it's in ACRID.

Observation class captures all the information for the execution of a given process. Lineage information is better served inside the metadata.

Stations fall into 1 of 3 sets: 1. sufficient information to calculate the sd when we want. 2. Not enough information internally but can calculate normals from data from elsewhere. 3. Can't get external information to calculate normals and sds.

Better if normals are recalculated if possible, rather than relying on static lists.